



NRL/FR/7180--10-10,196

## A Report of the NRL Technical Metrics Workshop 2008

JOSETTE P. FABRE

*Acoustic Simulation, Measurements, and Tactics Branch*

*Acoustics Division*

April 9, 2010

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE (DD-MM-YYYY) 09-04-2010		2. REPORT TYPE Formal Report		3. DATES COVERED (From - To) January 2007 to September 2008	
4. TITLE AND SUBTITLE  A Report of the NRL Technical Metrics Workshop 2008				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Josette P. Fabre				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 6031	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Naval Research Laboratory Stennis Space Center, MS 39529-5004				8. PERFORMING ORGANIZATION REPORT NUMBER  NRL/FR/7180--10-10,196	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  Naval Research Laboratory 4555 Overlook Avenue, SW Washington, DC 20375-5320				10. SPONSOR / MONITOR'S ACRONYM(S)	
				11. SPONSOR / MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT  A Technical Metrics Workshop was developed by the NRL Acoustics Division and held at NRL Stennis Space Center, MS, from April 29 to May 1, 2008. The primary goal of the workshop was to identify state-of-the-art technical and scientific metrics needed by the acoustic and oceanographic research and development communities. Secondary goals were to relate scientific/technical metrics to engineering and operational performance metrics, and to outline the propagation of errors and uncertainties using metrics variability. Thirty-eight presentations were given, representing participation from four U.S. Navy commands, six universities, and four private sector corporations, with foreign participation from the UK.  This report summarizes the presentations, discussions, and conclusions of the workshop. It will serve as a useful reference for building consistent technical metrics within the ocean and acoustic research and development communities, as well as providing common guidance for measuring exit and milestone criteria and for quantifying the contributions of interdisciplinary projects.					
15. SUBJECT TERMS Technical metrics   Scientific metrics   Antisubmarine warfare   Metrics   Acoustics   Geoacoustics   Oceanography					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT  Unlimited	18. NUMBER OF PAGES  317	19a. NAME OF RESPONSIBLE PERSON Josette Paquin Fabre
a. REPORT Unlimited	b. ABSTRACT Unlimited	c. THIS PAGE Unlimited			19b. TELEPHONE NUMBER (include area code) (228) 688-4879

## CONTENTS

EXECUTIVE SUMMARY .....	E-1
1. INTRODUCTION .....	1
2. TECHNICAL METRICS MOTIVATION .....	3
2.1 Commander, Naval Meteorology and Oceanography Command (CNMOC), N8 and N9 .....	3
2.2 Naval Oceanographic Office (NAVO) Oceanography .....	4
2.3 NAVO Acoustics .....	5
2.4 The Role of Scientific/Technical Metrics in an Operational METOC Metrics Program.....	6
2.5 The Use of Technical and Other Metrics within an Organization .....	7
3. STATE-OF-THE-ART TECHNICAL METRICS .....	7
3.1 Oceanography and Meteorology .....	7
3.2 State-of-the-Art METOC/Acoustic/Bottom Metrics.....	10
3.3 State-of-the-Art Uncertainty .....	16
3.4 Relating Technical Metrics to Management Level Metrics .....	20
4. DISCUSSION AND RECOMMENDATIONS .....	24
4.1 Definition of Technical and Scientific Metrics.....	24
4.2 Conclusions.....	25
4.3 The Way Ahead .....	28
ACKNOWLEDGMENTS .....	30
APPENDIX – NRL Technical Metrics 2008 Workshop Briefs.....	A-1

**NOTE: The appendix is in a separate file on this CD.**

## **EXECUTIVE SUMMARY**

Scientific/technical metrics are widely used by various communities at multiple levels from basic scientific analysis to decision making. The Navy has specific requirements that could benefit directly from such metrics. The overarching goal of this workshop, held at NRL Stennis from April 28 to May 1, 2008, was to have an interdisciplinary discussion on how these ongoing efforts could meet the higher level needs. Under this approach, more specific goals of the workshop were then to describe and identify technical and scientific metrics, build on current methods and discuss how these tools could be adapted/applied to operational and management metrics in the future.

During the discussion sessions, participants shared thoughts on how their efforts are intended to be used to contribute to higher level metrics such as those described in the provided material. Relevant inputs were collected during the workshop discussions and presentations. This report, resulting from the workshop, outlines the guidelines to the envisaged end-users, proposing the methods for the inclusion of technical, scientific and performance metrics into benchmarking and multi-criteria analysis of identified operational products and CONOPS.

The workshop emphasized how direct observations and model estimates can be combined to provide better operational guidance and assist in solving specific end user requirements. This will allow attainment of enhanced quantitative confidence for the end user, and therefore, provide not just an improved “final answer” but also identify scenario-dependent performance drivers and provide estimates of uncertainty that can more reliably and robustly inform operational decisions.

Mature metrics exist in each of the categories of metrics that are appropriate to this community: scientific/technical, performance and operational. The community has impressive metrics capabilities within the science and engineering domains. Similarly well-defined operations metrics have been developed within the Navy operations research community. What is generally missing is a general-purpose approach for tracing scientific improvements (e.g., a better temperature and salinity forecast) to engineering impacts (e.g., resulting improved ability to estimate SQS-53C detection ranges) to warfighting impacts (e.g., resulting improved ASW localization ranges resulting in the ability to meet MCO-2 warfighting objectives faster, with fewer resources, etc.). This methodology must provide the means to also trace uncertainties and errors from METOC data collection, assimilation, and modeling to end user operational effectiveness: correcting this shortfall is a primary long-term objective of the NRL Technical Metrics Committee (NTMC).

It was determined that the metric transmission loss (TL) difference and/or figure of merit (FOM) difference with enough information for uncertainty or sensitivity will provide a common scientific/technical assessment that can be computed at the output of each process and can then be easily translated to performance quantities.

The Navy acoustics S&T community is comfortable with TL, but much less so with FOM because of the inclusion of equipment capabilities, operator proficiency, and other factors which are hard to quantify in a scientifically rigorous fashion. Nevertheless, TL is of limited value operationally without an estimate of a FOM value or distribution of possible FOM values. This represents a disconnect between the operational and the Navy acoustics communities.

A general-purpose approach for tracing scientific improvements that have been expressed in terms of TL or FOM, to engineering impacts to warfighting will be developed and made available to the community. This methodology will be capable of tracing uncertainties and errors from METOC data collection, assimilation, and modeling to end user operational effectiveness. It must be as simple as possible so as to be relevant to multiple applications, with the knowledge that further analysis may be required. This effort must be coordinated with the existing capabilities on both ends (e.g., N81/N84/CNMOC and NAVO/NRL/R&D community) so as to provide consistent and agreed upon results. Some capabilities do currently exist, but are likely not in a format that can be easily used by the S&T community.

The next step in this metrics process will then be to research, identify and propose an approach or approaches for development of the aforementioned methodology. This approach will be different for various systems, but the initial focus will be on the current ASW systems discussed during the workshop.

Another Technical Metrics Workshop is tentatively planned for FY10. The purpose of that workshop will be threefold: 1. To present new technical metrics and progress on existing technical and related metrics since the 2008 workshop; 2. To present and refine the general procedure for deriving operational metrics from technical metrics; document the issues involved; and potentially to begin applying the procedure to a test case; and 3. To get feedback from the various entities involved on the technical metrics way ahead.

The NRL Technical Metrics workshop committee 2008 was lead by Josette P. Fabre, NRL SSC Code 7180 (acoustics). Other members of the committee included (in alphabetical order) Emanuel Coelho, NRL SSC Code 7320 (oceanography) / University of Southern Mississippi (USM), James Dykes, NRL SSC Code 7320 (oceanography), Pat Gallacher, NRL SSC Code 7330 (oceanography), Roger Gauss, NRL Code 7140 (acoustics), Dr. Joe Metzger, NRL SSC Code 7320 (oceanography) and Dr. Tom Murphree, NPS (meteorology/ oceanography/ climate/metrics).

The agenda was organized as follows. During the introduction or motivation, the shortfalls of military requirements that could benefit from applying some type of metrics analyses were presented. Next, the state-of-the-art metrics were presented, emphasizing how efforts could be steered towards the relevant problems. The state-of-the-art talks were categorized into the subject areas of METOC, acoustics, bottom and uncertainty. Finally, technical metrics were related to higher level decision making, addressing the questions on how to assimilate metrics of different types, along with other approaches to address the shortfalls identified at the beginning of the workshop. These efforts should identify our current technical / scientific metrics state, identify shortfalls and research directions and help design an end-to-end roadmap for the future. The next sections of the document summarize the talks; the briefs that were presented are given in the appendices. Each presenter contributed heavily to this portion of the document. The final section of this document provides the conclusions and recommendations of the workshop from the point of view of the committee members.

## A REPORT OF THE NRL TECHNICAL METRICS WORKSHOP 2008

### 1. INTRODUCTION

Technical and scientific metrics (Fig. 1) for the METOC and acoustics communities are as variant and diverse as are those who develop and use them. In a broad sense, they can be anything that is used to characterize the operational environment and its impact on performance. They can be narrowed down to more specific quantities based on the required applications, estimates of confidence in them, or by how they have changed over the time or area of interest. Gaps in them can be filled by fusing and assimilating data into models that analyze and forecast the environment. Metrics are further focused by computing performance estimates. At that point, evaluation of how an operation will be conducted or modified based on today's environment or forecast can be made. Finally, there are metrics that determine the accuracy of these estimates, how well we're doing and what is the return on the investment made in the tools that provide this environmental capability. **Underlying all of this is a tradeoff of accuracy versus efficiency.**

In order to make these metrics useful to the Fleet and to the high level decision makers, quantities that communicate things such as the accuracy, confidence, quality, efficiency, and impact must be developed and communicated in an understandable format.

The "pointy end" of Fig. 1 (i.e., the end that goes off the figure, because it is not in the technical or scientific realm) represents the high level metrics that provide the information for the "commanders" to help make decisions in both operations and budgets.

Ongoing operational metrics programs (e.g., those of Dr. Tom Murphree) that provide some examples as to how high level metrics that can be used to determine performance and impacts. Technical metrics will feed all of these operational metrics and the goal of this workshop is to understand these high level needs and to define which technical and scientific parameters feed which operational metrics for each application.

To that end, the goals of this workshop are to plan and start building "bridges" between the scientific/technical metrics community to the higher order decision and operational metrics communities. By discussing and understanding the high level needs, the scientific and technical (or research and development, R&D) community will be able to focus research in a way that will have a more direct and measurable impact. By discussing and understanding the existing research we can begin to develop guidance for existing and future efforts to provide capabilities that can be used by the decision makers. This is a difficult task, but worth the effort.

Section 2 presents the motivation for the report, and Section 3 presents the summaries of the state-of-the-art technical metrics efforts that were presented at the workshop and the beginnings of bridging the gap between the technical and operational communities. Section 4 provides discussion and recommendations from the workshop, and the Appendix (in a separate file on this disk) contains the workshop presentations.

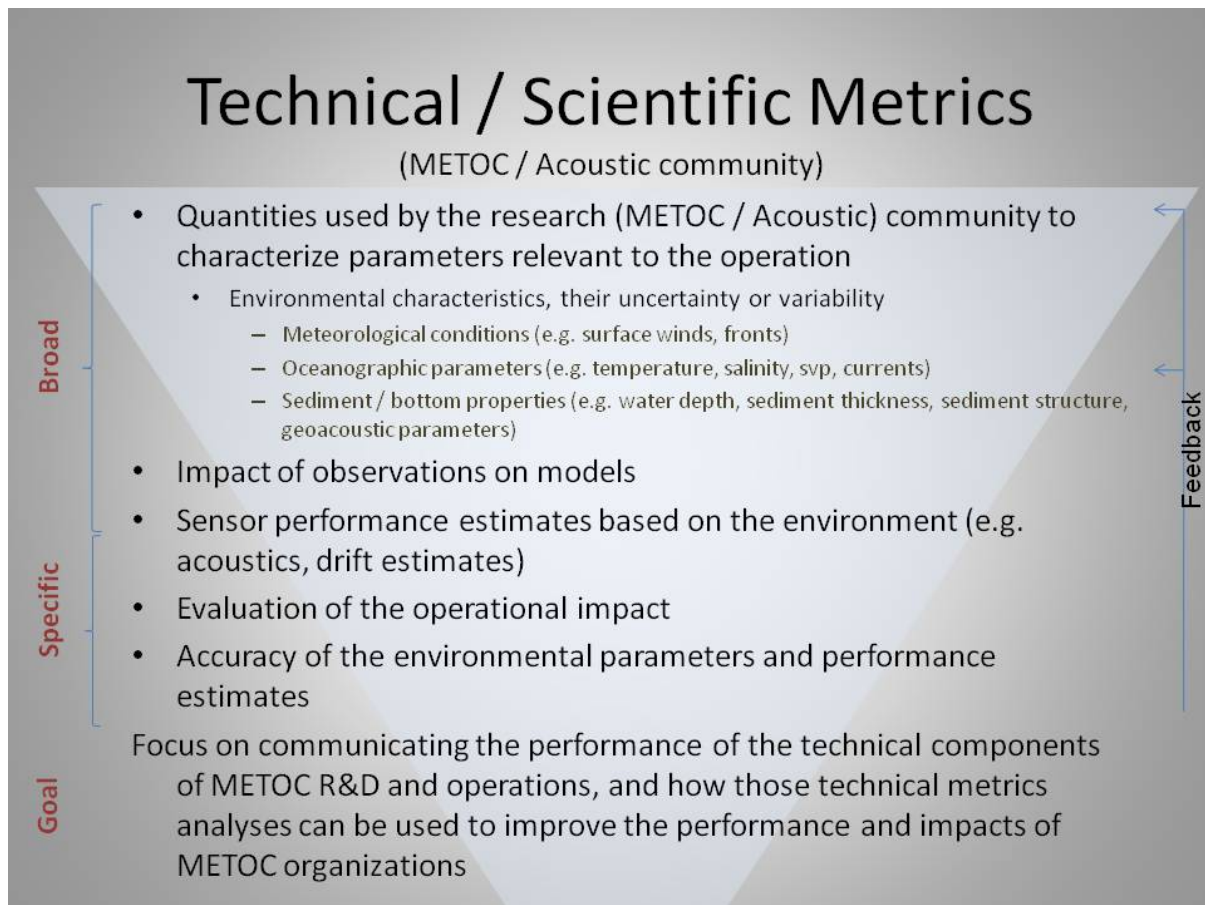


Fig. 1 — Technical/scientific metrics for the METOC/acoustic community

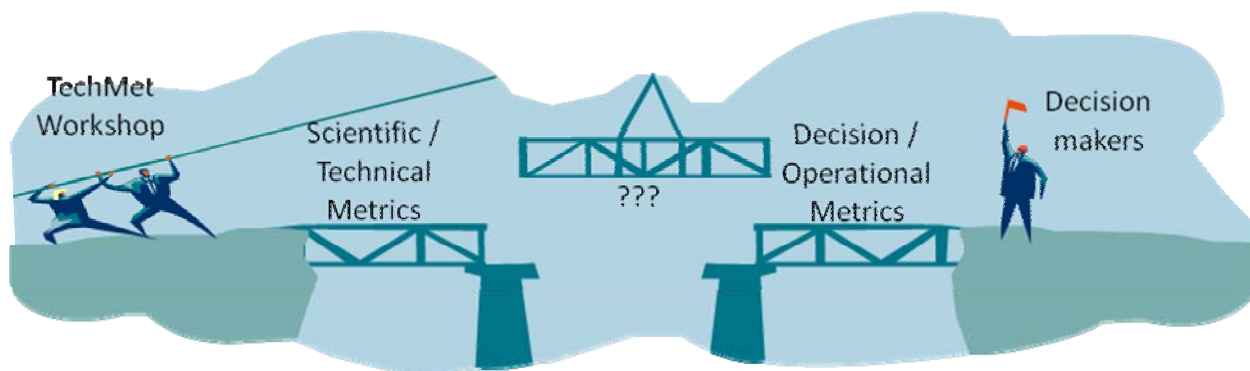


Fig. 2 — Building the bridge between the technical/scientific metrics and the decision-level metrics

## 2. TECHNICAL METRICS MOTIVATION

### 2.1 Commander, Naval Meteorology and Oceanography Command (CNMOC), N8 and N9

Dr. Merrill Stevens, Naval Meteorology and Oceanography Command (NMOC) Requirements, Programs and Assessments Department (N8) discussed metrics used by the NMOC staff to build readiness information for the current Five-Year Defense Plan (FYDP) submission. The metrics used by the NMOC N8 staff are primarily to support the Department of Defense (DoD) Planning, Programming, Budgeting, and Execution System (PPBES). The NMOC Technology Transition and Integration (N9) staff uses metrics such as research, development, test and evaluation (RDT&E) appropriations (i.e., 6.1-6.7), technology readiness levels (TRLs) (i.e., 1 through 9), and Joint Capabilities Integration and Development System (JCIDS) milestones and phases, from the concept development phase to production and fielding. Other organizations within the meteorology and oceanography (METOC) community use various types of metrics, from the *scientific* and *technical* metrics used by the science and technology (S&T) organizations to the *operational* metrics used by the warfighting units. Linking these metrics would show a more direct line-of-sight and the relationship between these metrics (see Fig. 3). This linkage would allow sharing, or reuse, of metrics, where metrics used by developers could be shared (reused) by operators and decision-makers; thereby making better use of the metrics and ultimately improving the efficiency and effectiveness of the programs, processes and products associated with the metrics. A better understanding of each METOC community and their metrics will help facilitate such linkage. Soon the Navy will start reporting domain-wide readiness metrics in the DoD Defense Readiness Reporting System, so a common understanding of that system will be required as well.

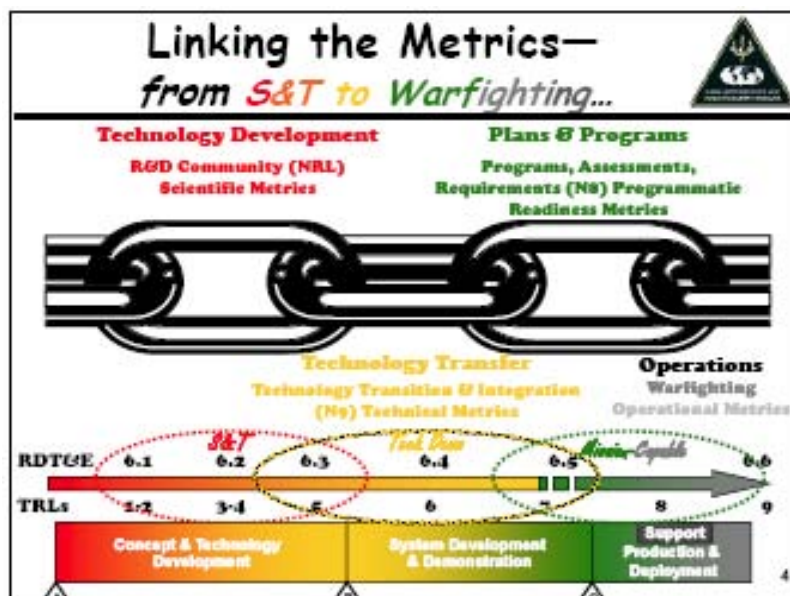


Fig. 3 — The link between scientific/technical metrics and warfighting metrics



Several examples of N8 readiness metrics were given. One example is under the category of Major Combat Operations (MCO) Strategic Intelligence Preparation of the Environment (IPE), depicting the number of T-AGS ships required to keep up with oceanographic survey requirements. Tier 2 metrics of the MCO IPE category included the associated operating and sustainment costs such as ship charter and hire, salaries of the crew, travel costs, equipment repairs and supply refurbishments. Other examples were given (Appendix A). There are gradations (sometimes called *tiers*) of these metrics along with risk information. It is easy to see how the metrics can get complicated at this level and as you “drill down” to the technical level, the problem becomes arduous. At the Fleet and OPNAV level, metrics are based on performance, cost, readiness, risk, etc.

Dr. Stevens proposed a sample metrics schema that defined different levels of metrics, to include scientific, technical, operational and organizational metrics, and gave an example of the linkage between these four categories of metrics using ocean gliders. It would be very useful to show the linkage between issues being addressed by the scientific community and their metrics to the higher level, operational and organizational metrics. More discussion of this and the committee’s suggestions is provided in the conclusions.

## 2.2 Naval Oceanographic Office (NAVO) Oceanography

Dennis Krynen (NAVO) presented the state of the ocean prediction products and interpretation of those products by the fleet and by acoustic modelers. Within NAVO, as an operational center; the oceanography division runs operational models that provide analyses and forecasts of parameters such as temperature, salinity, and currents, as well as derived parameters such as sound speed and sonic layer depth (SLD). They also run wave models for height direction, period and surf. Their models run every day and their products must go out in a timely fashion, and in formats ready for inclusion in tactical decision aids (TDAs). Seven days a week, 8 hours a day, ocean forecasters are part of the product preparation process, providing information such as the quality of the delivered products. There are requirements for ocean products for many warfare areas, for example, SPECWAR requires temperature and currents, but the highest is Anti-Submarine Warfare (ASW) support, and is the focus of his brief. The primary recipients of these products are the Naval Oceanography ASW Teams (NOATS) and the NAVO Acoustic division in addition to the operational customers. A big focus is on the product confidence and how to put that accuracy, confidence or uncertainty detail in terms that can be used by the customers. Currently, this information is provided in general terms, such as high, medium and low confidence based on the data that are assimilated into the models. Another major concern is the significant computation time of the NAVO Acoustic Performance Surface product. The timing of the updates of this product is considered, for example ASW parameters (e.g., SLD and layer gradients) and their spatial and temporal dynamics are used to help determine the performance surface ocean input parameters.

Their main tool is the High Resolution Navy Coastal Ocean Model (HI-NCOM) which provides ocean analyses and forecasts at 1/36° horizontal spatial resolution and 40 layers in depth. The ocean model quality assessments are currently very subjective, because 1) the model assimilates data, averages to the grid, uses inputs from FNMOC and 2) many assumptions are made in the process. It is not a simple task to consider all these and other factors in a model quality assessment. Additionally, this new HI-NCOM capability offers much more spatial and temporal data than ever before, increasing the challenge to process and analyze all the information and translate it to the Fleet and other users in a timely fashion.

NAVO makes a distinction was made between metrics and measures. They need metrics that provide information on for example, where to best place assets. In their opinion, metrics help them make a decision and measures are what they use to estimate the metrics. A metric might be the impact of an observation (BT or glider measurement) on the model validity.

Ocean forecasters must provide products quickly; they need metrics that indicate accuracy of estimates that come from the models or data acquisition systems. Currently, for example, it is very difficult to determine the impact of a degraded wind prediction on modeled sound velocity. The error or confidence information needs to be propagated through to the products, e.g., performance surface, so that an understanding of the source of that error can be communicated along with the product. There is never enough data or time to properly evaluate the model. They are currently trying to collect model statistics over a long period of time (years) and compare to observations to determine the model performance. Currently, their primary confidence metrics are based on model to data comparisons and knowledge of historic oceanography, seasonal trends, etc.

An example NAVO product is the Tactical Ocean Forecast Analysis (TOFA) that provides a text description, relevant ocean features for the area and time frame of interest, monthly or seasonal trends, but there is currently no set “recipe” for the product, that is up to the forecaster.

Another issue regarding model quality assessment is feedback. NAVO often supports pre-exercise analyses, so there is rarely data for comparison in those situations. During exercises, operations change frequently and the data collected during these operations can be sparse. Some hindcast analysis is being done in the research community to assess the models’ capabilities.

## **2.3 NAVO Acoustics**

Keith Atkinson discussed the NAVO Acoustics Performance Surface product and the need for related metrics. The input to the performance surface is sound speed from HI-NCOM, and bathymetry and sediment descriptions from databases. They mainly use some derived acoustic parameters (primarily SLD) and bathymetry for determining the acoustic run set-up. Their primary concern, currently, is the number of products they can provide each day. They do not have the resources to run all that they would like so they need information (metrics) that would allow them to smartly determine appropriate forecast or analysis times for which to run the performance surface, and which acoustic scenarios, spatial resolutions, etc. to consider. They would like to have metrics that allow them to adapt the grid resolution to the environments and to determine when and if a propagation run must be updated. They would also like to provide some measures of the uncertainty of their products.

Before the performance surface is generated, they consider parameters such as surface duct and cut off frequencies and their variations over time and space. Their main focus is on the change metrics and sensitivities to various oceanographic parameters in order to determine when to run the performance surface. Change metrics are indicators of how much the environment has changed over a time frame or a depth, etc. Currently, the performance surface is generated using PC-IMAT and the Sonar Tactical Decision Aid (STDA) to create a grid and run narrow band transmission loss (TL) for radials at each grid point, then convert to signal excess (SE) based on a receiver operator curve (ROC) for the specified sonar and a uniformly distributed target. Then the radials are collapsed into a single value of probability of detection. The acousticians then analyze and tune the product to the scenario and generate a PowerPoint

brief. They are beginning a validation and verification (V&V) process to determine whether or not the performance surface reflects what the operators in the field are seeing, and to later trace back any disagreements to the appropriate parameters (e.g., ocean model, ambient noise, target specifications, etc.).

Because of the complexity of the performance surface and the many inputs to it, there are many sources of inaccuracy and uncertainty in the product. Each of these was discussed and is summarized in the briefing (Appendix A, page 5??). In summary, NAVO needs ways to rapidly determine when to update performance surface runs based on changing sound speed fields, optimize TL grids to the environment, determine confidence of and incorporate uncertainty into the performance surface in terms that the Fleet can understand. They need metrics that can allow them to 1) do product assessments, reconstruction and analysis (R&A) and V&V and 2) understand the sensitivities of the performance surface to the various input parameters.

## **2.4 The Role of Scientific/Technical Metrics in an Operational METOC Metrics Program**

Dr. Tom Murphree presented a brief to help develop the framework for thinking about scientific/technical metrics. He provided some basic definitions and concepts (see slides Appendix A). There are many levels of metrics that can be provided to the customers (the warfighters), not just what is measured, but what is calculated, derived from models, and the impacts of products on missions. One way to think about metrics such as SLD is to ask why there is interest in this quantity. There is interest because the outcome of the operation being supported will be partially impacted by that quantity and how well that quantity is represented. So the real concern is performance. Everything discussed so far falls into METOC performance metrics. To really understand the significance of the product, customer performance metrics are required, for example, did the customer accomplish their goal, and did they accomplish it safely? The desire is to understand the connection between the support products and the customer's success. If the METOC performance metrics and the customer performance metrics are combined, operational impact metrics can be defined. Scientific/technical metrics are a subcategory of METOC performance metrics, which is the performance of technical systems to generate end user products.

Scientific/technical metrics can help answer METOC metrics questions such as what are the gaps in METOC support, which METOC products are worth generating, is there a more efficient way to produce these products, what is the uncertainty in our products and how much confidence should we or our customers have in our products? Many of the answers will depend on thresholds and other factors. Metrics can get overwhelming very quickly and we need to be careful with defining the thresholds by which we make the decisions.

In conclusion, the scientific/technical metrics are a fundamental part of a METOC metrics effort. For scientific/technical metrics to be most effective, they should be developed and used with understanding of end user thresholds; with understanding of the sensitivities of and uncertainties in the end user planning, outcomes and costs; in close coordination with the development and use of operational impacts metrics and so that they are well aligned with the overall goals and the practical applications of the organizations metrics program. Suggested focus topics for scientific/technical metrics are the identification of physical factors to/for which end use planning and outcomes are most sensitive/ uncertain. Thresholds are not necessarily the best indicators of end user sensitivity/uncertainty.

## 2.5 The Use of Technical and Other Metrics within an Organization

Steve Woll (CDR USN (ret)), Weatherflow, Inc. gave a brief providing a high level perspective of metrics. There are three trends, first, metrics are here to stay so they need to be used to the best advantage. Metrics are now used across many areas including business, baseball performers, troops in Iraq, and are often briefed to the public. Second, there are always perceived failures in the government. Lately, things are worse, the American public feels that things are not going well, the culture of performance is metrics. The third trend is funding shortfalls, technology is more expensive, and there is less discretionary funding. Most decisions are being made based on resources. The people making these decisions are high level people with limited technical background. Additionally we are flooded with information operationally and tactically and there are many sources of information. If things are too complex, they will be ignored, there will never be enough time, money or people and decisions will be made regardless of the level of information, with metrics being used to make those decisions.

People are comfortable with what they know and expect. Decisions that reinforce pre-existing beliefs are easier to sell than those that challenge conventional wisdom. Metrics should be correct and not “cherry picked”, i.e., they should be in terms the recipient understands but referenced to terms that reflect the technical issues. Conclusions should be easily drawn from the metrics. Finally, they should be simple; if they have to be explained to a significant degree, they are too complicated.

## 3. STATE-OF-THE-ART TECHNICAL METRICS

Here, state-of-the-art metrics are presented from the viewpoint of the scientific performers. We present this in two categories (oceanography and meteorology) and acoustic.

### 3.1 Oceanography and Meteorology

#### 3.1.1 ASW METOC Metrics from Valiant Shield 07

Bruce Ford, Clear Science, Inc. presented results from a feasibility study in ASW operational impact metrics conducted during Valiant Shield 07 (VS07). Real time, operational data were collected to discern the impact of meteorology and oceanographic information on Naval operating forces. This information will be used not only to assess impact on the warfighter during VS07, but also to improve data collection ideas/methods.

Data were collected primarily through direct sampling using on-scene data collectors. Data were collected from multiple warfare areas such as surface forces and maritime patrol aircraft. Data collectors assembled observations, forecasts, warfighter plans, outcomes and impressions of customers, as well as recommendations offered by NOATs. The resulting data were summarized and are provided Appendix A.

An apparent error evaluation of COAMPS and NCOM fields was also conducted. This preliminary analysis showed that COAMPS forecasted winds that were in error due to model failure to sufficiently capture timing, location and/or intensity of known types of atmospheric patterns and processes (e.g., trade winds, monsoons, deep convection, and low-level cyclonic circulations). The NCOM lessons learned included: apparent errors did not have an obvious correlation to the corresponding atmospheric errors; errors with small spatial structure appeared to be associated with in-situ observation locations; error with large spatial scale may be related to small scale errors; errors developed during the assimilations at early

forecast times persisted through the 48 hour forecast; and errors reset with each new run did not persist from one run to the next. Preliminary findings and questions are detailed Appendix A.

The data collection effort for VS07 represents the most ambitious real-time data collection effort to date directed toward metrics calculation exclusively. The operational impact/accuracy metrics presented may represent proxy metrics for acoustic analysis and ocean model validation metrics; thus this effort is extremely important to the technical metrics community.

The examples of data collected for VS07 originate from a project in its early stages and additional, more ambitious data collection efforts are planned as this project moves toward the goal of continuously collecting, computing and displaying ASW related operational impact and other types of metrics.

### *3.1.2 New Operationally Relevant Scientific Metrics for Evaluation of Spatial Predictions*

Barbara Brown of the National Center for Atmospheric Research (NCAR) presented a brief explaining that metrics are important issues in atmospheric science as well as in oceanographic and atmospheric research. A paper published recently in the Bulletin of the American Meteorological Society addressed the issue of metrics from putting the forecast together to the end product including value added or taken away at each step in the process. New methods in atmospheric science focus on spatial coherence in order to quantify the operational or user relevance. In the traditional approach the focus has been on precipitation or convection, and the skill depends on the application. The problems with this traditional approach is that it does not indicate what was right or wrong with the forecast and this approach is ultra-sensitive to errors in simulation of local phenomena. The goal is to come up with alternate approaches. Spatial forecast verification techniques aim to account for uncertainties in timing and location, account for spatial structure, provide information on error in physical terms and provide information that is diagnostic and meaningful to the users. Neighborhood methods are object and feature based methods that give credit to “close” forecasts and help determine if anything is gained by higher resolution. The Method for Object-based Diagnostic Evaluation (MODE) measures forecast attributes that are of interest to users, it mimics how a human would identify storms and evaluates forecasts. Issues identified include ensuring that the metrics do not change the forecast or diagnostic, the forecast/diagnostic should represent the “true” expectation of what presumably will happen. One must identify attributes that are relevant to particular applications and can be related to “customer effectiveness.” In conclusion, these spatial methods could be applied to other areas, such as oceanography.

### *3.1.3 Model Verification*

Dr. Gregg Jacobs, NRL 7320 (Oceanography) presented metrics from two points of view, the forward problem and the inverse problem. The scope and the context of the metrics must be set from end-to-end, i.e., from METOC systems to mission impact. The mission impact can be thought of as a vector that has mappings between warfare mission impact, physical information, physical processes, system processes and the full system descriptions. The forward motivation is determining the effect of an upstream (environmental system) perturbation on the downstream (mission impact) consequence. The issues are where in the system to measure the perturbation, how to test the hypothesized solution, what are the sensitivities of the various components, what meaningful metrics exist, and the final impact. The inverse motivation is given a desired result, what should be changed, i.e., what is the sensitivity to the physical

input, what are the error levels, and where should resources be invested. Having metrics as the long term goal, we must examine the entire scope of the problem. These metrics not only link the beginning (mission impact) to the end (environmental systems), but also the links between each step along the path. The forward motivation approach takes into account the language and understanding of the warfighters and decision makers, whilst the inverse look translates the needs at the pointy end of the spear into actions that oceanographer must take to improve the product. Still, much of the information required to develop these metrics is unknown.

An example was given that illustrated determining the need for a satellite altimeter in ocean modeling, determining how many altimeters were required, what the impact of the altimeter is on the ocean model, how accurate it is, and what its orbit must be. Several experiments assimilating different altimeter measurements provided variations of the expected errors of sea surface height (SSH) from MODAS 2D (input to NLOM). Using a correlation of the accuracy of the ocean model to the probability of detection map that was generated for the RIMPAC08 area of interest (AOI), effects of altimeters measurements can be seen. With a 100-dB cutoff, it can be shown in plots of RMS and maximas how the accuracy of the synthetic profiles can decrease the overall error. In addition, a decrease in transmission loss error can be correlated to the added time needed to search for targets.

An analysis methodology was devised to determine the best path from start to finish according to performance predictions, and to then determine the actual performance based on the path selected. The available environmental predictions determined the performance predictions. Signal excess was improved with increased altimeter observations. In the end, ship loss prediction error was decreased significantly (as high as 70%) with the introduction of altimeter measurements.

### *3.1.4 Metrics Used to Evaluate, Validate, and Transition the Global HYCOM / NCODA / PIPS System*

Joe Metzger, NRL 7323 (Oceanography) presented a survey of the types of metrics used in the validation of a global ocean nowcast/forecast system, namely the 1/12° HYbrid Coordinate Ocean Model (HYCOM)/Navy Coupled Ocean Data Assimilation (NCODA) system which is scheduled for delivery and transition to NAVO at the end of FY08. It will eventually replace the 1/8° Navy Coastal Ocean Model (Global NCOM) but only after a thorough validation has shown that it “adds value” over the existing products. The metrics used by the research community can be different from those used by the operational community, and so NRL and NAVO came to a consensus on a set of validation tasks to be performed. Because of the size and scope of the effort, the entire process will span a couple of years. A brief synopsis of these validation tasks follows.

To first order, any global ocean model must accurately reproduce the large scale circulation features, e.g., the basin-wide gyre systems and western boundary currents. By qualitatively and quantitatively comparing observed and simulated SSH, the large scale can be evaluated. The variability of the oceanic mesoscale is another measure of system realism, i.e., whether the meandering fronts and eddies are properly simulated. Comparisons of satellite derived SSH variability or observation based eddy kinetic energy aid in determining if the model has realistic level and distribution of energy. Of utmost importance to the operational community is the system’s ability to nowcast and forecast (out to at least 5 days) the 3-D temperature and salinity structure and ASW-related fields such as the mixed layer depth, sonic layer depth, deep sound channel axis, and below layer gradient. Such quantities can be derived from observed

profiles and compared against simulated results. Accurate nowcasts and forecasts of sea surface temperature (SST) can be validated against many different data types, including satellite MCSST, fixed or drifting buoys, and ship observations. One of the key functions of a global ocean nowcast/forecast system is the provision of boundary conditions to regional and coastal models. The same types of validation described above can be applied to a nested inner model that has been forced with boundary conditions from two different outer models, thus determining the impact of the boundary forcing.

### *3.1.5 Internal Waves and Internal Tides*

Dr. Tim Duda, Woods Hole Oceanographic Institute, presented work done in collaboration with Dr. Pat Gallacher (NRL 7331, Oceanography), which suggests technical metrics to quantify the areas and times when large amplitude nonlinear internal waves and internal tides will be generated and where they will propagate. Internal Tides and nonlinear internal waves that develop from them are the cause of sound-speed anomalies in stratified continental shelf waters. In contrast to surface (barotropic) tides, internal tides are not well predictable at this time. Internal tides are known to result from the surface tides interacting with sloping bathymetry. They are also known to evolve into packets of steep, short-wavelength nonlinear internal waves a significant fraction of the time. Usable prediction of internal tides (wavelengths in excess of 30 km) would require extreme resolution, in part because they are sensitive to bathymetric details. Usable prediction of nonlinear internal waves would require the internal tide prediction, and may also require non-hydrostatic modeling. In lieu of precise prediction of internal tides, it may be possible to combine bathymetric slopes, stratification properties, and surface tide transports into a time-and space varying predictor (metric) for internal tide activity in both the semidiurnal and diurnal tidal bands.

Internal tides and nonlinear internal waves that develop from them are the cause of sound-speed anomalies in stratified continental shelf waters. In contrast to surface (barotropic) tides, internal tides are not well predictable at this time. Internal tides are known to result from the surface tides interacting with sloping bathymetry. They are also known to evolve into packets of steep, short-wavelength nonlinear internal waves a significant fraction of the time. More energetic internal tides would be expected to form nonlinear waves more rapidly than weaker ones. Usable prediction of internal tides (wavelengths in excess of 30 km) and their energies with computational regional ocean models would require extreme resolution, in part because the generation of these waves is sensitive to bathymetric details. Usable prediction of nonlinear internal waves would require internal tide prediction as a prerequisite, and may also require non-hydrostatic modeling. The process of barotropic wave to baroclinic wave energy conversion requires strong forcing that coherently drives waves.

## **3.2 State-of-the-Art METOC/Acoustic/Bottom Metrics**

### *3.2.1 Relationship between Horizontal-Array Performance and Internal-Wave Spectra*

Dr. Peter Mignerey of NRL Acoustics, presented work done in how horizontal coherence relates to internal waves. Decision metrics often used to describe characteristics are the receiver operating characteristic (ROC), which depend on array gain (AG). This type of metric is at the output of the beamformer. For this application, the data are at the hydrophone level, where there is a large signal to noise ratio (SNR) that allows more accurate estimate of coherence. AG depends on acoustic coherence, which depends on modal coherence and finally on the internal wave spectrum. The exponential coherence

and phase structure functions are presented. The coherence function is not well known in shallow water, so the power law is used based on array gain. Carey (Boston University) showed estimates based on work done around the world with coherence lengths of  $100 \lambda$  for deep water and  $30 \lambda$  for shallow water. The acoustic power can be related to modal power by examining the cross-modal amplitude matrix.

Dr. Mignerey has developed a horizontal coherence function that relates the environment to the coherence, by developing a modal phase-structure function, which relates the internal wave and acoustic wave parameters. In summary, the array gain depends on the coherence function and the horizontal coherence depends on the internal wave spectrum (energy and spectral shape). This coherence theory remains a work in progress, but comparisons to the empirical curves show promise.

### *3.2.2 Data/Model Comparison for Littoral Soliton Packets and their Impact on Array Performance Using Integrated 3-D Ocean-Acoustic Modeling*

Dr. Roger Oba of NRL acoustics presented a proof of concept for metrics to evaluate array performance degradation due to soliton-induced acoustic variability. The metric validation of the integrated model focuses initially on the model's capability to predict the salient features of the internal wave-induced sound speed variations and to quantify the consequent degradation of array beamforming performance. To evaluate those capabilities, array performance measures are applied to both the observed and modeled beamformed acoustic output. The data-model comparison provides a proof of concept for the integrated model as follows. Initial conditions derived from ASIAEx oceanographic and archival data are used in NMCO, a 3-D, sub-mesoscale hydrodynamic computational model, to simulate the evolution and propagation of internal soliton packets in variable bathymetry. The modeled soliton packet evolution is shown to be comparable to that in the ASIAEx data, and includes observed features such as wave front curvature due to shoaling. The modeled temperature and salinity distributions are mapped to time-evolving sound speed fields for input to a 3-D acoustic field propagation code. The modeled and observed array performances are compared. Quantitative measures of model/data array performance, including array signal gain and bearing accuracy, show significantly improved predictive capability over modeling using only hydrostatic models with tide or 2-D modeling.

### *3.2.3 Acoustic Boundary- and Fish-Interaction Metrics for Active ASW Sonar Performance Predictions*

Dr. Roger Gauss of NRL Acoustics presented his work that showed the environment plays an integral role in the performance of any active sonar system where reverberation from the ocean boundaries and fish can mask desired signals or create false targets. The influence of these environmental scatterers can vary greatly depending both on the local oceanography, geology, and biology, and on the sonar characteristics and geometry. This talk begins with a high-level overview of the crucial undersea scattering phenomena shown or predicted to impact ASW sonar performance in both deep and shallow water. While separable to some degree in deep-water environments, bottom, surface and fish scattering are not so cleanly separable in shallow-water environments. The dominant scattering or propagation mechanisms must often be inferred by repeated broadband measurements, coupled with comparisons to physics-based model predictions. Furthermore, each shallow-water environment can be unique; techniques that work in one environment do not necessarily extrapolate to another (stressing the need for in-situ measurements).



This talk then reviewed state-of-the-art metrics of single-bounce acoustic boundary and fish interactions for active ASW sonar applications, including boundary loss (BL), scattering strength (SS), frequency shifts, and the probability of false alarm (PFA). All can be sensitive to environmental (boundary and biologic conditions) and system (scattering angle, frequency) parameters. BL models are key to accurately predicting long-range propagation and reverberation, especially in shallow-water environments where multiple bounce geometries occur. (Both surface and bottom losses can be important, e.g., interactions with the rough sea surface can scatter the acoustic energy into the seafloor above the critical angle where it is rapidly lost, so that at long range mainly low-angle energy remains.) SS is important to all reverberation calculations, with bottom SS (BSS) typically the most important. The empirical Lambert's Law is the default BSS model used in reverberation models; while fast to compute, it cannot capture physics or extrapolate in either frequency or geometry (especially bistatically). While physics-based BSS models exist, required geophysical inputs are often lacking (such as bottom roughness, which can be important below the critical angle, and so, for long-range acoustics). At the other boundary, efficient physics-based surface SS (SSS) models exist that rely environmentally on only the wind speed; hence applicable in near real time and on regional scales. Fish scattering is generally well understood and modeled; however, biological inputs for the acoustic models are usually lacking, especially in shallow water where fish can display high degrees of spatial and temporal variability. Modeling frequency shifts is important to developing/assessing low-Doppler detection schemes. Physical models are available capable of predicting the mean frequency-shift characteristics of acoustic signals (< 5 kHz) scattered from the moving sea surface, bubble clouds and fish, with the dominant frequency shifts being at the Bragg lines for air-water interface backscatter, and at zero shift (with a Gaussian distribution of shifts about this peak) for bubble cloud and fish backscatter. Statistically, scattering amplitude distributions (PDFs) of normalized reverberation data typically exhibit non-Rayleigh behavior for discrete scatterers (seafloor heterogeneities, bubble clouds, fish) leading to appreciable PFAs. Physical models (K and Poisson-Rayleigh distributions) that estimate the number of discrete scatterers per unit area exist, but have very limited validation by field data (where both the composition of the scatterers and their spatiotemporal distributions are rarely known). For all, a series of data-model comparisons demonstrated the importance of using physics-based tools to predict the acoustic boundary- and fish-interaction responses.

### 3.2.4 Active System Metrics: Arrival Structure

Dr. James Fulford of NRL Acoustics presented work on evaluating the arrival structure for an active system. Active system performance involves knowledge of active target returns, active noise or reverberation, and the ambient noise environment. Heuristically, the time series arising from an active system can be written as:

$$S(t) = S_d(t) + S_T(t) + S_v(t) + S_b(t) + S_s(t) + S_a(t),$$

where  $d$  refers to the direct path time series,  $T$  refers to the target time series,  $v$  refers to the volume reverberation time series,  $b$  refers to the bottom reverberation time series,  $s$  refers to the surface reverberation time series, and  $a$  is the ambient noise time series. The performance model then predicts the ratio of the target time series to non-target time series as seen through a specified detector. Metrics of active system performance prediction usually involve one of two concepts: (1) comparison of predicted detection parameters (usually signal excess and range) with measured values or (2) comparison of a

prediction of the subset of the system performance prediction with a measurement of the subset. A detection parameter based system metric is related to acceptance testing – a system exists, it meets some criteria and will be accepted or it will be rejected. A subsystem testing metric is a tool for model development, and ideally identifying the components of the system performance model where technical innovation would result in performance prediction improvement.

Analysis of the structure of the active source dependent parts of the active time series reveals that each of the terms has an explicit dependency on source/receive beam patterns, and the arrival structure of the acoustic energy. For illustrative purpose the direct path term of the active series time series will be examined. The direct path term can be written as:

$$S_d(t) = \sum_{i=1}^n I(t - t_i) B_s(\theta_i, \lambda_i) L_i B_R(\theta_j, \lambda_j)$$

where the paths,  $L_i$ , are those that propagate directly from source (defined by source function  $I$ ) to the receiver involving only forward scattering.  $B_s(\theta_i, \lambda_i)$  is the source beam pattern in the vertical direction ( $\theta$ ), and the azimuthal direction ( $\lambda$ ), for the  $i^{\text{th}}$  source path.  $B_R(\theta_j, \lambda_j)$  is the receiver beam pattern in the vertical direction ( $\theta$ ) and the azimuthal direction ( $\lambda$ ), for the  $j^{\text{th}}$  receiver path. The direct path is in effect the transient direct response of the receiver to source activation. This equation is similar to the equation for the response of a transient transmission for a passive system. The other terms in the active system time series involve scattering from a target, a boundary, or a volume element. In theory the source and receive beam patterns are known, the source function is known, and the environment is known. Computing the paths, and applying the relationship between the components of the direct path will lead to a predicted direct path time series. This calculation should be repeatable at any point in the computational domain; in general data are measured in a few locations at best, so all points where there are measurements must be considered in constructing the metric. The metric that is sought shows need to show both the temporal structure of the arrival, and the energy. Thus two calculations are made: the correlation coefficient  $r$ , and the amplitude difference  $z$  between the observed time series  $O$ , and the predicted  $S$  as follows:

$$r = \frac{(n \sum S_i O_i - \sum S_i \sum O_i)}{\left\{ \left( n \sum S_i^2 - (\sum S_i)^2 \right) \left( n \sum O_i^2 - (\sum O_i)^2 \right) \right\}^{-1/2}}$$

$$z = \max \{ |S_i - O_i| \}$$

These two results are used to calculate a metric at each location

$$m = \left\{ 1 - r^2 \right\}^+ \frac{z}{\max |O_i|}$$

The total metric for the prediction is the maximum  $m$  which occurs in the prediction space.

It must be noted (Robert Miyamoto, APL/UW) that in an operational scenario, the data necessary for this analysis may not be available.

### 3.2.5 Metrics to Evaluate Acoustic Predictions

Ms. Josette Fabre presented work on comparing single frequency, single source transmission loss predictions as well as area coverage. Transmission loss (TL) can be computed using several methods, the most common of which are rays and Gaussian beams for mid to high frequencies, normal modes for low frequencies, and parabolic equation methods for low to mid-frequencies. TL models can vary significantly in how they compute the field, their range and depth resolutions can vary, their grid types can vary (triangular, rectangular) and the source description varies. There are a number of ways to compare TL. Line plots and field plots can be displayed and visually compared; models can be computed on similar grids or range averaged to compute direct differences; weighted differences or area differences can be computed. A general rule is that at a 3-dB (corresponding to a doubling of intensity) is an acceptable TL difference. Metrics for comparing TL vary based on the application of the TL estimate, and there are currently no documented, consistent metrics published for these comparisons, though many comparison methods have been documented as they pertain to various applications. An example was shown for a mid-frequency case where TL was predicted using both a high resolution (comparable to a measured profile) sound speed profile and a smoothed profile (comparable to a modeled profile). TL comparisons were done using the direct output of the TL model and the range averaged TL. Statistics can also be computed with depth in order to maintain the vertical structure of the field. Mean differences for all ranges and mean magnitude differences for all ranges were computed. This type of comparison can be done when the TL curve needs to “match” at all ranges. Difference at the maximum range is computed for cases where the TL will be used only at the range of interest. Area and area coverage differences were computed for a sensor coverage type metric. Additionally, when comparing two TL curves the high values of loss, can be de-weighted as very high values of loss (120 to 140 dB) rarely contribute to the acoustic field and high differences can occur that do not affect most products of TL. The output of the TL model is generally acoustic pressure or intensity and comparisons can be made at that point, however, the values are very large and can exaggerate differences. The comparison results varied based on the application and this emphasizes the need for different metrics for different applications.

Next, acoustic data were considered. Transmission loss data are collected during scientific exercises that support research and development (R&D) and also for geoacoustic surveys of sediments. TL is rarely, if ever, collected by the Fleet. Ambient noise is often collected, and it is a very complicated function of TL. Detection ranges are often documented by the Fleet and that information can be used to verify TL models. There are many issues involved in comparing models to data. First, there are different ways to express the environmental inputs, and we can rarely consider a fully realistic environment, (e.g., internal waves, resolutions on the scale of acoustic wavelengths). Assumptions are made by the acoustic models to make the problem solvable, and assumptions are made by the environmental models for the same reason. Data are often flawed as well due to the sensors that take the data, the recorders, the calibration process, etc. Measured data are generally compared using the same techniques as for modeled data.

Uncertainty can be estimated using both measured and modeled data. There are several ways to estimate uncertainty in the models. Currently, two methods are being employed by NRL SSC, first, ensembles of oceanography are being generated (Bishop, Rowley, and Coelho 2007) and TL is computed using those estimates. Second, Zingarelli (2008) has developed a method of computing the uncertainty at the output of the TL model that considers modal variation over TL input parameters such as sound speed,

water depth, source and receiver depths and bandwidth. Examples of both techniques were shown, both methods have limitations, but are very good estimates until the very difficult solution of acoustic uncertainty can be solved.

### *3.2.6 Acoustic Performance Metrics*

Mr. Steven Dennis of NRL Acoustics discussed various acoustic performance metrics. Transmission loss (TL) is a quantitative measure of the weakening of sound travelling between two points. TL is considered to be the sum of loss due to spreading and loss due to attenuation; it is the basis for most acoustic performance metrics. Signal excess (SE) is the received signal (in dB) in excess of that required for detection. The figure of merit (FOM) for passive sonar is defined as the maximum TL for which a signal will be detected some percentage of the time. Probability conditions are implied in the term recognition differential (RD). SE is a fundamental detection metric. The detection range is the ranges throughout which the system will be able to detect a target, and the maximum detection range is the maximum range from the source at which the target can be detected. Coverage is the area throughout which a sensor is able to detect a target. Coverage is computed for geographic sectors around a source (or sources) as an area (range and azimuth) for which the sensor at that location would be able to detect a target, i.e., there is positive signal excess. Another measure, related to coverage is visibility, the percentage of coverage grid locations that are able to detect another given location. Once the coverage area at every location in an area of interest is computed, the sensor placement can be assessed by looking at the largest coverage areas. Optimal sensor paths can be computed by following the shortest path between the highest covered areas. Similarly, visibility maps can be used to avoid detection. Several metrics are used to evaluate search paths, the percentage of targets detected, the time to complete the search and the validity of the predicted route.

The concept of integrated acoustic coverage (IAC) (Fabre 2007) was introduced. Coverage can be integrated over all possible source configurations, (depth, frequency, etc.) to obtain a better performance estimate. IAC can also be computed over various times or ensembles in order to provide estimates of uncertainty.

The advantages of using acoustic coverage as a metric are, it reduces a large amount of range and azimuth dependent sensor performance information into one easily interpreted value; it can be visualized in the form of a coverage map which gives an easily readable overview of the sensor's performance in that environment; calculation of coverage automatically gives visibility information for vulnerability applications; it is a straightforward and quick calculation; it is expressed in units of area and can thus be easily manipulated mathematically; and it can be used to represent the effects of environmental uncertainty on acoustic sensor performance.

### *3.2.7 Acoustic Predictions vs Real Life - A Sound Idea*

Dr. Robert Miyamoto presented his work on comparing acoustic predictions to measurements.

Acoustic performance predictions are generally made without regard to a systematic approach to evaluating whether or not we are making accurate predictions. An evaluation of prediction tools for active acoustic, multi-static, sonobuoy data using exercise reconstruction has proven useful for evaluating the acoustic performance prediction. Estimates of signal-to-noise using post-flight mission data in

performance prediction tools were compared to measured, signal-to-noise of validated detections. The evaluation of these comparisons demonstrated shortcomings in performance predictions that has led to improvements in performance prediction accuracy.

### *3.2.8 A New Project to Quantify Bottom Loss Anisotropy in Ocean Bottom Sediments*

Dr. Warren Wood presented work being undertaken in the next fiscal year.

Beginning in FY09 the Naval Research Laboratory is funding a study of the anisotropy of ocean bottom sediments; particularly the anisotropy associated with finescale faulting that is pervasive on continental slopes. Anisotropy of sound speed and attenuation in deep-sea sediments can cause errors in calculation of seafloor reflection coefficients and predictions of bottom loss exceeding a factor of five; this is particularly important near critical angles (that can shift 10° or more) and at small grazing-angles. Our specific objective is to measure seismic anisotropy in unconsolidated, deep-water sediments to determine its impact on bottom loss and bottom scatter. This objective includes investigating compressional wave-speed anisotropy (which has been quantified to some extent in deeper, consolidated sediments) and attenuation anisotropy (which is poorly known in all sediments). Existing numerical models of anisotropic wave propagation will be used to predict wave-speed and attenuation anisotropy at all grazing angles for potential field sites. Subsequent field measurements will be acquired using combined deep-towed and bottom mounted systems in order to cover all grazing angles; the data will be analyzed to determine observable anisotropy as well as to obtain bottom loss. The products will be algorithms and system design recommendations that will provide transition recipients with a measurement capability for populating bottom loss and/or geoacoustic databases in the 200 Hz to 4 kHz band.

## **3.3 State-of-the-Art Uncertainty**

### *3.3.1 ONR Quantifying, Predicting, and Exploiting Uncertainty DRI – Program Overview and Science Plan*

Dr. Pat Cross gave a discussion of the geography and key environmental factors, oceanography and acoustic science objectives, and the basics of the science plan to address the objectives, including the array of sensors (ocean and acoustic) that will be used and a quick overview of the field experiment schedule for this ONR program. The purpose of this program is to try to ascertain the factors that drive acoustic uncertainty. This is a 5-year program, beginning with a pilot study in September, and a main cruise in 2009. This effort represents a collaboration between the U.S. and Taiwan for improvement of performance prediction and performance. At the end of the effort, they would like to be able to exploit uncertainty. They are operating in the Okinawa trough and plan to estimate end-to-end uncertainty of systems of interest. They will transmit frequent pings to make robust statistical curves and compare that to operator calls (reality). Another metric will be to generate predicted probability of detection and compare against human performance.

### *3.3.2 Managing Uncertainty: Ocean Nowcasting and Forecasting Issues*

Emanuel Coelho of the University of Southern Mississippi and the Naval Research Laboratory Oceanography Division presented his work on ocean modeling.

In the Battlespace on Demand (BOND) framework, environmental information from operational centers can be used to assist and reduce the risk of operational decisions, by working at three different levels or tiers. On a first tier, environmental data are measured or estimated using forecast and analysis models, on a second tier environmental data are used as input into performance models to estimate performance layers, cost functions and thresholds related with operational systems efficiency; and finally at the third level tier 1 and 2 derived products and their uncertainty are integrated to assist the decision process into operational and tactical planning and tasking, that use specific concepts of operations (CONOPS) and rules of engagement (RoE) (Fig. 4).

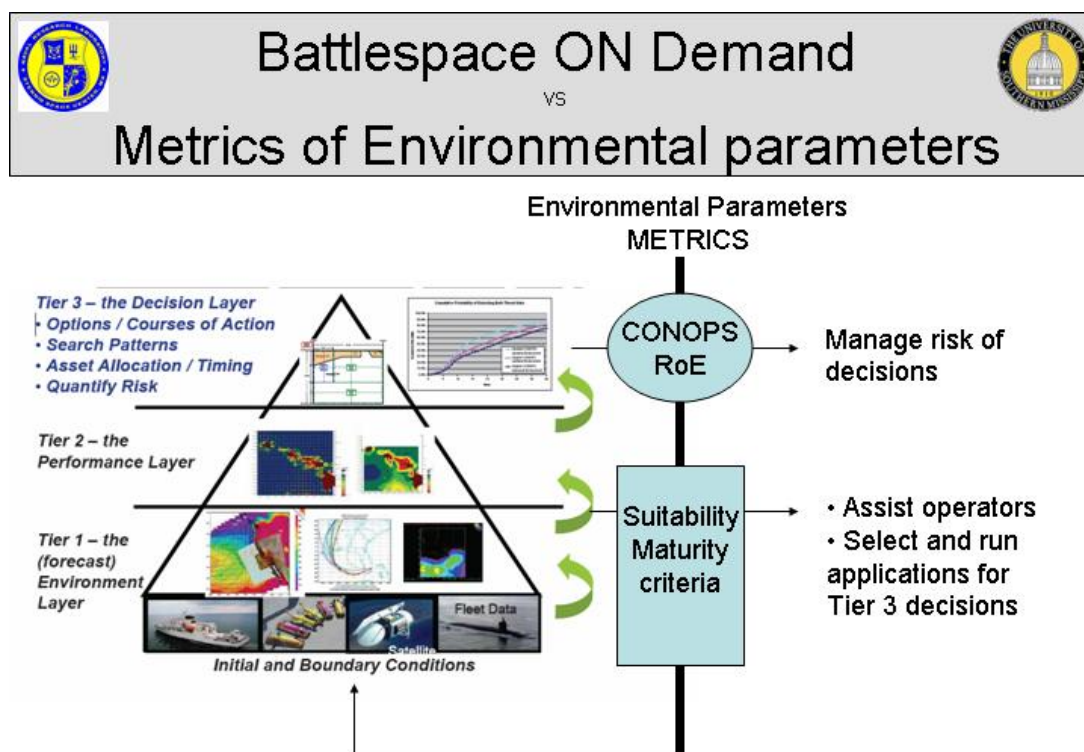


Fig. 4 – Battlespace On Demand (BOnD) concept vs metrics of environmental parameters

Besides common data and information management procedures dealing with the required resources, data flow and timings, each level or tier has different issues that can be addressed using tailored metrics approaches:

- At tier 3, the primary challenge will be to manage the risk of decisions related to the stochastic nature of the environment e.g., using metrics to benchmark tier 1 and 2 products using CONOPS

and RoE designed criteria such that one could assess the risk of making the wrong operational or tactical decisions because of misleading or incomplete environmental information.

- At tiers 1 and 2, the challenge will be to evaluate inputs and outputs in terms of their suitability and maturity for the required applications, i.e., using metrics to benchmark and estimate bias and confidence intervals of the analysis and predicted input variables and estimate how they are converted into bias and confidence intervals of the environmental thresholds, cost functions and systems efficiency-performance.

One can then conclude tiers 1 to 3 metrics are also required because of the stochastic nature of the environmental variables estimates. These metrics should assist not only the decision process but also feedback into the operational centers such they could be used to fine-tune procedures and products towards their end applications (end-to-end approach).

When addressing the predictability of these variables and associated metrics, multiple sources of errors need to be considered. They are associated with the initialization and boundary conditions of models, numerical approximations, modeling strategies, observation representation errors and unresolved scales. Furthermore, when nesting or coupling models of different kind, these errors will have multiple cross-correlations, creating a very complex non-linear multi-scale, interdisciplinary problem that we can call Uncertainty Cascade (UC).

By its nature, the UC is better characterized through coupled observations (e.g., combining acoustic and hydrographic data for sound speed profile estimation) and models (e.g., combining waves, currents and bathymetry for nearshore dynamics prediction), and requires extensive collection of multi-scale interdisciplinary local data and model outputs. These facts motivate the need to simplify the procedures by using multiple criteria analysis (MCA) based on overarching self-consistent metrics systems.

Approaches of MCA using Monte-Carlo simulations to design tier 1 to tier 3 metrics have been tested using scientific trial data and recent naval exercises for surface drift problems and simple ASW scenarios and showed that there are results mature enough and worth being included in compatible methodologies and tested for possible near-future transitions.

### *3.3.3 Accounting for Uncertainty in Simulation-based Prediction for Ocean-Acoustics Modeling*

Dr. Steve Finette presented some of his ONR-sponsored work in acoustic uncertainty.

If a metric is defined for the purpose of model-data comparison, the modeled (simulated) component should account for uncertainty or errors (just like the measured component should) in order to interpret the comparison in an objective, quantitative manner. In practice, one is always faced with incomplete environmental information when simulating oceanographic and acoustic field properties in an ocean waveguide. The validity of simulation based prediction schemes depends on the assumption that either all environmental information necessary for the solution of the problem is known or, if this information is only partially available, that the resulting uncertainty in one's knowledge of the environment can be objectively quantified and included in the result. If neither of these conditions is met, the conclusions or decisions that are based on the prediction are of questionable validity, as are any metric dependent conclusions linked to such a prediction. We are currently investigating a probability-based methodology

for incorporating environmental uncertainty into both oceanographic and ocean-acoustic modeling for the purpose of quantifying the uncertainty in simulation-based numerical prediction schemes. Here, the environment is treated in a probabilistic manner, as a function of random variables or fields in order to quantify incomplete information in the result of a numerical simulation. Probability density functions describing uncertainty in the resulting oceanographic or acoustic properties could then be used in defining metrics of interest.

### *3.3.4 Adaptive Sampling Oceanographic and Acoustic Cost Functions as Metrics for Model Validation*

Kevin Heaney presented his work on acoustic metrics for adaptive sampling applications.

An adaptive sampling approach based upon the non-linear optimization of a user-defined set of cost-function has been developed to determine the “optimal” placement of ocean sampling sensor systems. The goal is to sample the environment in regions which bring the ocean model the closest to the true ocean. The key questions are: “What defines the term closest?”, and “What is optimal to the user?”. Several candidate oceanographic and acoustic constituent cost functions have been developed which can be summed (via a normalized weighted linear sum) to generate the overall cost-functions. The definitions of these cost functions provide useful metrics for evaluating the difference between the model ocean and the true ocean. Oceanographic cost functions, which can lead to oceanographic metrics are locations of the fronts and model RMS variability of the Tsigma. Acoustic cost functions include acoustic sensitivity (coherent TL, incoherent TL and mode amplitude correlations) across the ensemble of ocean forecasts. With the definition of these metrics, the quantitative evaluation of the value added by adaptive sampling and data assimilation can be determined.

### *3.3.5 Algorithm for Bathymetry Fusion with Uncertainty Assessment*

Dr. Paul Elmore of NRL Mapping, Charting, and Geodesy division presented work that he and Dr. Chad Steed performed.

We discuss findings of our recent literature review of current fusion techniques used for bathymetry or other geospatial data, as motivated by the Naval Oceanographic Office’s need for new intelligent fusion algorithms - combining two or more data sets in a manner that accounts for data uncertainty - for gridded and in-situ bathymetric data sets. Based on this review, the most robust published approach for building new bathymetry fusion algorithms uses both Loess interpolation to obtain a trend surface, followed by Kriging of residuals to recapture finer details lost from smoothing. In addition, if in-situ soundings are used, Monte Carlo simulations are run to estimate depth error induced by position errors. The technique also provides the means to liberally estimate errors for navigation safety. This talk reviews this approach and discusses plans to build, validate, and transition the algorithm to the Naval Oceanographic Office for use with future bathymetry databases.

### *3.3.6 Fitting Data, but Poor Predictions: Reverberation Prediction Uncertainty When Seabed Parameters are Derived from Reverberation Measurements*

Dr. Roger Gauss presented a brief prepared by Dr. Charles Holland of Pennsylvania State University, Applied Research Laboratory.



For many decades, researchers have been developing inverse techniques for estimating seabed parameters from reverberation data, notably scattering strength. Generally, the angular dependence of the scattering kernel is unknown and is either solved for or assumed fixed. In either case, agreement is typically quite good between the measured reverberation and that modeled (by fitting scattering parameters). However, what are the resulting uncertainties in a reverberation prediction if the ocean or geometry changes? The main results of the paper are that 1) these prediction uncertainties are surprisingly large, of order 10 dB at 10 km and thus 2) traditional/current methods for reverberation inversion should be augmented, mitigating the large prediction uncertainties by an additional measurement. Several options for additional measurements are discussed.

### **3.4 Relating Technical Metrics to Management Level Metrics**

#### *3.4.1 Bridging Techniques/Best Practices for METOC Impacts Metrics Data Collection*

Bruce Ford of Clear Science, Inc. presented work conducted by himself, Dr. Tom Murphree, and David Meyer of the Naval Postgraduate School, Paul Vodola, Matt McNamara, Luke Piepkorn, and Ed Weitzner of Systems Planning and Analysis (SPA), and Dr. Bob Miyamoto of APL-UW.

In order to assess the impact of CNMOC/NAVO models and other products on military effectiveness, technical/scientific metrics systems must be bridged to METOC impacts data collection systems that can measure (1) the overall effectiveness of the METOC support provided and (2) the effectiveness of the warfighter.

The collection of data for determining the impacts of METOC products on military operations is generally problematic. In some special situations, data collection may be completely automated. But in the vast majority of cases, data collection in part by humans is required. Data collection systems that adhere to a growing set of best practices increase the likelihood of collecting accurate, continual, quantitative, and objective data. Such best practices fall into three basic categories:

1. Institutionalization within the military unit – addressing paradigm shifts that must occur to ensure regular and consistent data collection
2. Human behavioral factors – understanding the priorities and limitations of those tasked with entering critical data
3. Human-machine interface design – designing a system that is intuitive and allows rapid entry, updating, administration, and use by the military unit's managers.

This presentation proposes a set of best practices for use in building METOC impacts data collection systems. Examples of operational impacts metrics systems that we have developed for USAF and USN units were presented, along with examples of the resulting impacts metrics.

Process data were proposed as a bridge between the scientific/technical metrics and decision/operational metrics. Process data include what is produced, how many, how often, how accurate, how efficient, etc.

METOC-related process metrics guidance would need to apply to processes in general, provide a common framework for METOC-related process metrics, apply regardless of the class of metrics

(scientific, technical, operational), attempt to establish a common metrics language, attempt to prioritize potential metrics (if we cannot collect everything we need, what are the highest priorities?), become the expectation for emerging processes (institutionalization), include cross-process utility (when one process may not have knowledge of the other) and finally, one size will not fit all.

The scope of data collection must be focused at the correct level. Very often leaders will want grand-scale metrics that depend on smaller scale metrics (without a system providing the smaller scale metrics). Metrics are most useful when they provide information to multiple levels of the organization e.g., individual forecaster, immediate supervisor, forecast activity commander, directorate and higher. Fact-based metrics are most useful when developed from data from the lowest levels of the organization. It is critical to collect data on the smallest “unit” of support (e.g., forecast, mitigation recommendation); quality higher level metrics (directorate, CNMOC) rely on lower level data collection/metrics.

### *3.4.2 Operational Ocean Forecasting Research and Development at the Met Office*

Dr. Ray Mahdon of the UK Met Office presented an overview of their ocean forecasting capability.

The Ocean Forecasting Research and Development (OFRD) group at the Met Office is responsible for development and maintenance of the operational ocean forecasting systems. REMIT covers short range ocean forecasting and does not include seasonal, interannual, or climate time scales. Their definition of “operational ocean forecasts” is routine forecasts suitable for use in critical activities and they must be: robust, with service level agreements and backup procedures; timely, to meet agreed delivery schedules; and supported with 24/7 operator cover and help desk. The key OFRD customers include the Royal Navy, Environmental Agency, Department of Environment, Food & Rural Affairs (DEFRA), Offshore Industry and the Public Weather Service.

Dr. Mahdon summarized their overall process, which includes data assimilation via the Forecasting Ocean Assimilation Model (FOAM), daily analyses and 5-day forecasts, hindcast capabilities, and wave modeling. FOAM configurations and data types were summarized, the Northwest European Shelf nested models were discussed and marine ecosystem modeling for water clarity and algal bloom warning were presented. POLCOMS (-ERSEM) MOD (Proudman Oceanographic Laboratory Community model system; European Regional Seas Ecosystem Model; Military of Defence) potential users and applications include: Currents and winds for SAR (synthetic aperture radar) and oil and mine drift; Water clarity products for autonomous underwater vehicles (AUVs), divers, submarines evasion/detection; Underwater acoustics; Currents for operations for mission planning tools and dispersion forecasts; TKE products; Currents and winds for oil drift; Nearshore waves for MOD amphibious operations; Relocatable model capability for Royal Navy; Inter-comparison and model validation with Navy in-situ observations (expendable bathythermographs (XBT), conductivity temperature depth (CTD), acoustic Doppler current profiler (ADCP) and sea-soar); generation of climatologies for Royal Navy; and Marine mammal distribution for acoustic sonar applications.

### *3.4.3 OPTEST Report East China Sea Navy Coastal Ocean Model (ECS-NCOM)*

Dr. Frank Bub of the Naval Oceanographic Office (NAVO) presented a brief he gave to the Administrative Model Oversight Panel (AMOP) that resulted in a recommendation to adopt the model and declare it operational. The operational test (OPTEST) objective was to demonstrate an advancement of capabilities over MODAS. These advancements consisted of 1) the application of physics vice statistics, 2) forecasts to 120 hours, 3) observation quality control with data assimilation using NCODA, and 4) increased numerical skill. This particular model set up covers mainly the East China Sea at high resolution. Forecasts out to 72 hours every 3 hours include temperature, salinity, currents, elevation, and derived RP33 acoustic properties for ASW. Much data were collected in September and October of 2007 and were assimilated via NCODA after the OPTEST comparisons were made. A thorough statistical approach included using Gaussian statistics and spectral analysis, the latter to help address internal tides or propagation of internal waves, which was difficult to model. The first three objectives were easily met, whilst there was incremental improvement as shown in the statistics.

### *3.4.4 Automation of Metrics Rapid Transition Program Effort*

James Dykes of the Naval Research Laboratory, Oceanography Division spoke about some SPAWAR/ONR funded work that is being transitioned to NAVO. Aspects of an automation of ocean product metrics to be transitioned to the ASW Reach-back Cell in NAVO NP1 include software tools to help in assessing ocean model and acoustics performance. In addition metrics of the performance of mission-impacting products are also to be produced. All these elements provide the means and information to evaluate the end-to-end system and apply many of the concepts presented in this workshop. Model skill is the primary driver in this Rapid Transition Program (RTP) and is addressed with automated data collection and statistics-generating software for the user's perusal on a graphical display. Tools in MATLAB at the oceanographers' and acousticians' disposal provided immediate information regarding model performance and their effects on acoustical statistics, aiding the user in providing useful analysis information and quality digital data to the ASW METOC support personnel who support the customers directly. A web based survey system provides a means to collect information regarding the use of the METOC products. Summaries are compiled and provided to the command and support personnel to help make decisions on improving METOC support. Also, included were slides showing snapshots of the ASW Reach-back Cell Operational Analysis System (ARCOAS), a GIS-based set of tools used to display the ocean model output and related performance statistics.

### *3.4.5 An End-to-end System Analysis*

Dr. Robert Miyamoto presented an example of a recent end-to-end systems analysis using an active acoustic multistatic sonobuoy system. The issues of defining such an end-to-end system are presented and the data required in order to evaluate such an end-to-end analysis are identified. A probabilistic approach to the evaluation of the end-to-system is presented using fleet exercise data.

### *3.4.6 Advanced Visualization Techniques for Undersea Warfare*

Chad Steed of NRL Code 7400 presented geovisual analytic techniques and how they can be applied to undersea warfare data. Analytics (statistics and artificial intelligence) and visualization techniques are applied to amplify cognition of undersea warfare data. There has been an unprecedented growth in the

quality and quantity of data in general, but also specifically of the environment for the Naval Oceanographic Office, NASA, and NOAA. These datasets hold great potential, but our ability to generate all these data is far outpacing our ability to understand them. Visualization is a key factor in coping with these data. The data can be reduced and refined by harnessing the high bandwidth human perceptual channel. The traditional visualization approach involves layered or separate plots. There are many problems with this approach, including change blindness, layer occlusion, and layer interface. Only a handful of layers can be displayed, and colormaps can lend emphasis to features that are not really the emphasis.

For undersea warfare, particularly for acoustics, the data are complex, multidimensional, and geospatially referenced. Also, the data are beginning to have associated uncertainty. It is challenging to display all this information in a comprehensible manner. New visualization techniques are being pursued based on human perception guidelines. The goal is to encode variables in a single display for analysis to avoid many perceptual issues. An example is shown below (from Healey and Walter 2001) (Fig. 5) where four variables are encoded into glyphs for weather data.

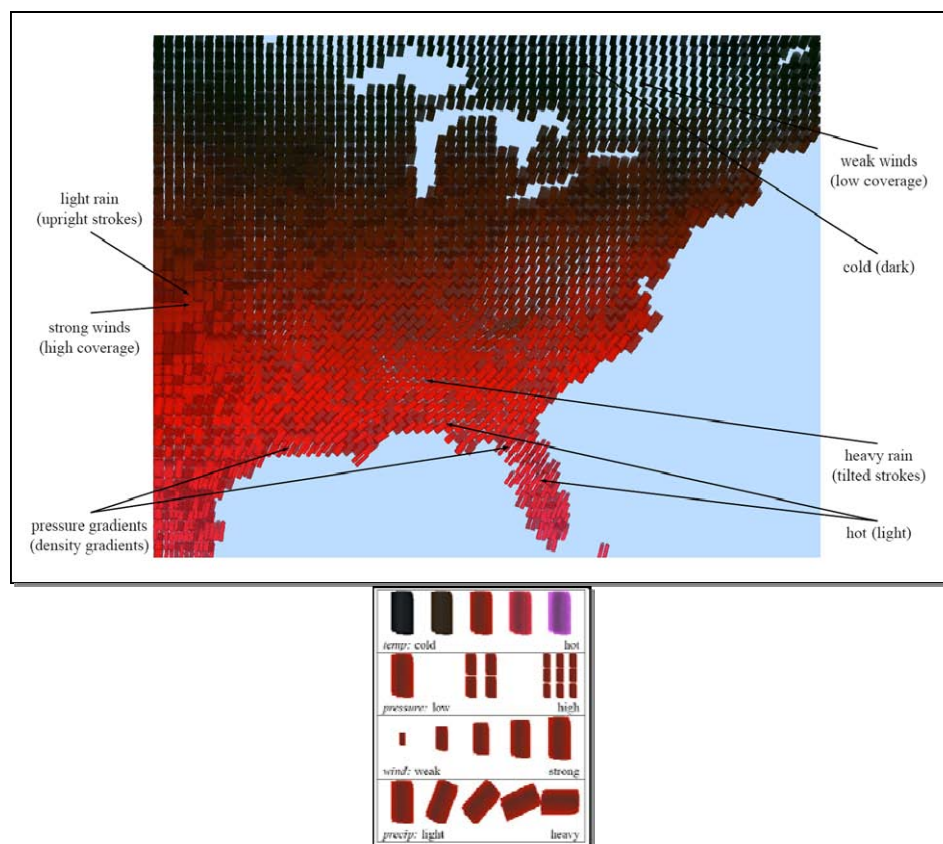


Fig. 5 – Example for weather visualization

Metrics are computed in the process of developing these advanced displays. These include computed thresholds for quick visual analysis (red, yellow, green), quantification of associations or correlations between data, and measures of clutter or density in the data. As these techniques advance, more metrics will be developed and defined.

#### *3.4.7 Deriving Metrics from a CNMOC Initiative on Reconstruction and Analysis*

Bruce Northridge of CNMOC described an initiative to develop a capability for reconstruction and analysis of Fleet contact data. It is very important to reconstruct and analyze acoustic exercise data to better understand the data as well as the quality of the METOC products and tools used to estimate the performance of the systems that are collecting the data, and to improve lessons learned. Several metrics are being developed and some were presented from an effort to reconstruct Fleet data compared to the NAVO acoustic performance surface predictions. These included distributions (spatial and geographic), statistics and ranges of contacts and opportunities, probabilities of detection vs number of contacts, as well as range, confusion matrices, and receiver operator curves derived from confusion matrices. In conclusion, the Navy spends a lot of money on ASW exercises and METOC products. These efforts will help improve lessons learned and knowledge of product quality. Many metrics have been developed that were derived from exercises and real-world events and the metrics will evolve and become more robust as their development continues.

## **4. DISCUSSION AND RECOMMENDATIONS**

At the beginning and end of each day, discussions were held based on topics identified during the day. These discussions were expanded by the committee and are included below, followed by committee recommendations based on the workshop.

### **4.1 Definition of Technical and Scientific Metrics**

It is necessary to define what is meant by technical and scientific metrics, as well as operational metrics so that there is consistency in metrics programs and discussions in this community. The group settled on the following definitions:

**Metric** – a metric is an agreed upon set of standard measures, or more generally, a system of parameters, or a set of ways of quantitatively and periodically measuring or assessing a process (scientific, technical, assets, operations, enterprise, etc.), along with the procedures to carry out measurements and the procedures for the interpretation of the assessment in the light of previous or comparable assessments.

- **Scientific/technical metric** – metrics of the performance of the scientific/technological systems from which final products (e.g., forecasts) are developed (e.g., performance metrics for sensor networks, data assimilation, modeling processing, model outputs, etc). Scientific metrics are metrics that describe how well we understand and can model a physical process. They capture the degree to which a physical property can be measured, predicted, and/or forecast. For example, how well can the ocean temperature and salinity field be nowcast or forecast over a grid of ocean locations and depths? Environmental spatial and temporal variability and the degree to

which the underlying physics can be modeled or represented in data are drivers of these scientific metrics. Scientific metrics can be physics based or empirical (data driven). They link different phases of each process toward the systems end.

- **Performance metric** – captures the degree to which warfighting effectiveness is impacted. For example, one performance metric might be the ability to estimate sonar detection range as a function of environmental state and levels of uncertainty. How well is the underlying physics understood and how does it translate to the performance prediction?
- **Operational metrics** – measure the aspects of the operation that address mission success, such as effectiveness and safe operating procedures. For example, METOC Performance Metrics are metrics of the success of a METOC organization in conducting its operations, for example, the success of its processes, products and service (e.g., accuracy, sensitivity, uncertainty of METOC forecasts of sonic layer depth).

## 4.2 Conclusions

The motivation for improved metrics is in response to the CNO N81/N84 need to show quantitative traceability from METOC model/database improvements to Navy warfighting impacts in scenarios of interest. All major DoD acquisition decisions must be justified with clear and convincing warfighting (impact) analysis. This analysis must explain in detail how the proposed METOC improvements will lead to improved warfighting effectiveness.

Mature metrics exist in each of the categories of metrics that are appropriate to this community: scientific/technical, performance, and operational. The community has impressive metrics capabilities within the science and engineering domains. Similarly well-defined operations metrics have been developed within the Navy operations research community. What is generally missing is a general-purpose approach for tracing scientific improvements (e.g., a better temperature and salinity forecast) to engineering impacts (e.g., resulting improved ability to estimate SQS-53C detection ranges) to warfighting impacts (e.g., resulting improved ASW localization ranges resulting in the ability to meet warfighting objectives faster, with fewer resources, etc.). This methodology must provide the means to also trace uncertainties and errors from METOC data collection, assimilation, and modeling to end-user operational effectiveness; correcting this shortfall is a primary long-term objective of the NRL Technical Metrics Committee (NTMC).

Several examples of successfully tracing scientific metrics to engineering metrics, and engineering metrics to operational metrics were, however, presented at the NTMW. Significant discussion was undertaken to identify the metrics of interest and the appropriate way ahead. Steve Lingsch (CNMOC) outlined a starting point by mapping the acoustic modeling inputs through to the end user. After much discussion, the diagram below (Fig. 7) was agreed upon by the group. The colors of each box refer to the tiers in the Battlespace on Demand (BoND) pyramid. Tier 1, shown here in blue, refers to the environmental description layer. This information can be obtained from various sensors, from various models, and can be processed and stored in databases. Tier 2, shown in purple, represents the performance

layer, which consists of estimates of, in this case, acoustic performance. Tier 3, shown in red, represents the decision layer, that is, the end-user of the METOC products. Additionally, there are aspects of this community that are represented in multiple tiers of the BoND pyramid (Fig. 6). In Fig. 7, items representing tiers 1 and 2 are colored in green, and all three tiers are colored in orange. For clarification purposes, a Navy system the AN/SQS-53C was selected for use as an example. The examples for this system are given in green text in each box. The dashed lines around each box indicate that there is uncertainty associated with this quantity that must be accounted for and carried through to each application. The lines between each box show the connection between each item. For example, bathymetry, which can be measured or modeled and is databased in the Navy's DBDBV, is used as an input to the oceanographic model (Hi-Res NCOM). This information is then fed to the acoustic models for prediction of transmission loss for the given acoustic system parameters, and the resulting information is fed to sonar performance prediction systems such as the NAVO Performance Surface. The performance surface information can then be used for planning and decision making. Feedback from each performance assessment can be used in a number of ways, including but not limited to reconstruction and analysis, sensitivity studies, and operational impacts assessment.

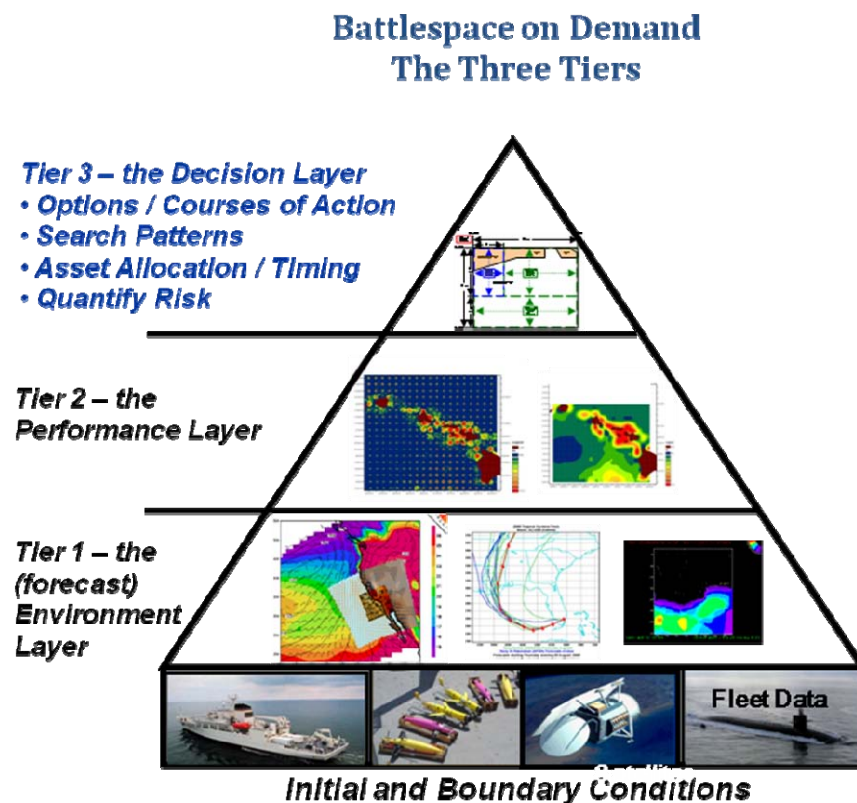


Fig. 6 – Battlespace on Demand pyramid (CNMOC 2007)

It was determined that the metric transmission loss (TL) difference and/or figure of merit (FOM) difference with enough information for uncertainty or sensitivity will provide a common scientific/technical assessment that can be computed at the output of each process and can then be easily translated to performance quantities.

The black box in the figure shows this metric. Fig. 8 shows a summary or overview of Fig. 7 without all the detail. The environmental inputs with uncertainty (tier 1) feed the acoustic models (tiers 1 and 2), which in turn feed the sonar performance model (tier 2). The performance model is used for decision making (tier 3) and in support of real-world events. Real-world events occur and can be used for assessment of the sonar performance model and the decision-making process and for reconstruction and analysis.

A very important concept is the sensitivity of the models to the environmental and other inputs. Quantifying sensitivities, in addition to uncertainty, is important for research direction and budgetary decisions.

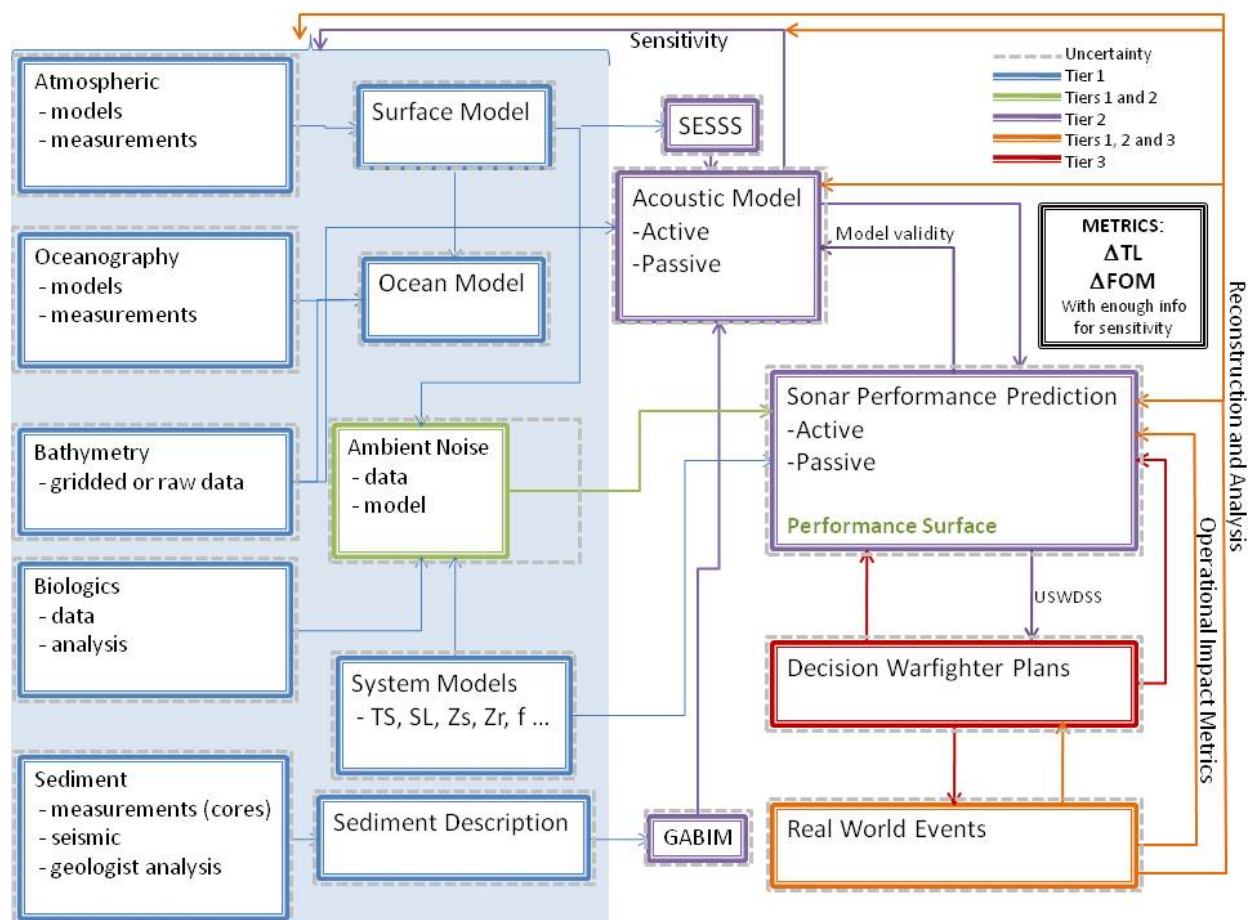


Fig. 7 – Technical metrics diagram. Examples for each category for a selected system are provided in green text in each box. The box colors identify the relationship to the BonD pyramid and the dashed line around each box indicates that uncertainty metrics should be considered. The uncertainty should also be translated between each category.



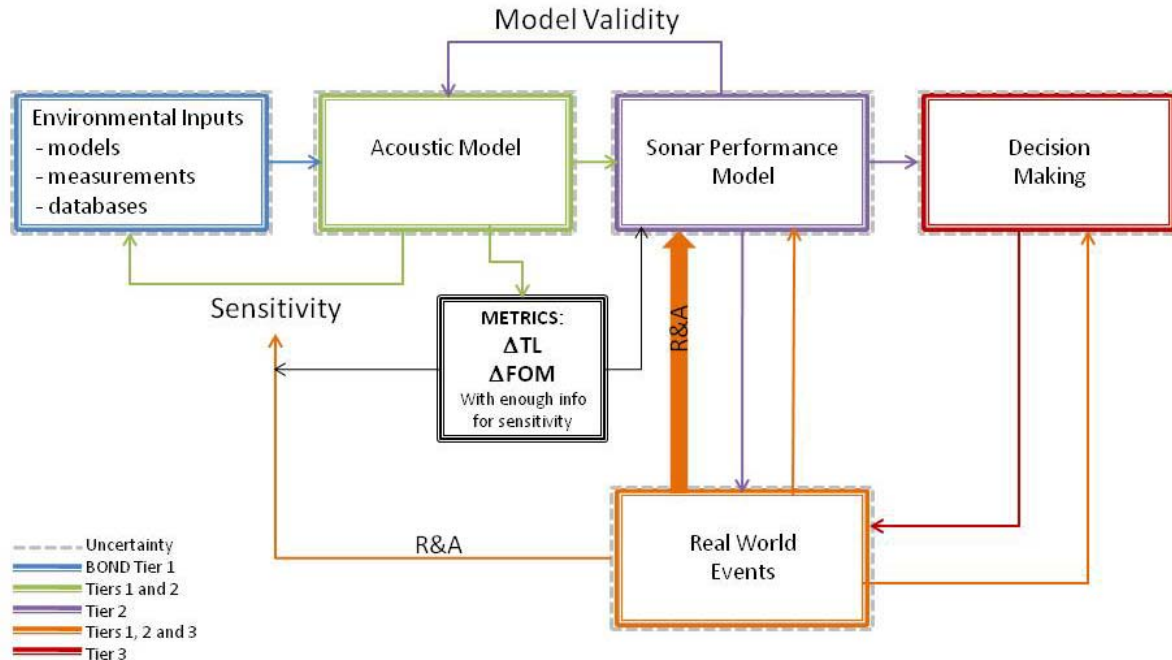


Fig. 8 – Technical metrics overview

### 4.3 The Way Ahead

The committee is pleased with the outcome of the workshop, though much work remains to be done. The METOC R&D community is now thinking along the lines of how their metrics need to be presented to the operational and decisional communities. A preliminary way ahead follows.

The decision-making community (e.g., N84) would like to quantify the value-added of METOC assimilation, modeling, database, or TDA developments. The operational community wants to know the same thing in addition to how to obtain and employ (e.g., CONOPs) new capabilities. Another important aspect of this is that for a new capability to be fielded, the resulting capability improvements need to be substantial enough to warrant the cost of transitioning to the new system.

The scientific community must be able to express their impacts in changes in transmission loss (TL) or figure of merit (FOM). For many, that requires running an acoustic model. Different models apply to different applications and each scientist should understand which models apply to their particular research application.

The Navy acoustics S&T community is comfortable with TL, but much less so with FOM because of the inclusion of other factors that are hard to quantify in a scientifically rigorous fashion. Nevertheless, TL is of limited value operationally without an estimate of a FOM value or distribution of possible FOM values. This represents a disconnect between the operational and the Navy acoustics communities.

Because of this, TDAs have been developed that try to deal with the FOM issue directly. Moving from TL to the other terms in the sonar equation could hence be part of a new focus on metrics.

Next, a general-purpose approach for tracing scientific improvements that have been expressed in terms of TL or FOM to engineering impacts to warfighting must be developed and made available to the community. This methodology must be capable of tracing uncertainties and errors from METOC data collection, assimilation, and modeling to end-user operational effectiveness. It must be as simple as possible so as to be relevant to multiple applications, with the knowledge that further analysis may be required. This effort must be coordinated with the existing capabilities on both ends (e.g., N81/N84/CNMOC and NAVO/NRL/R&D community) so as to provide consistent and agreed upon results. Some capabilities do currently exist, but are likely not in a format that can be easily used by the S&T community.

An ASW Impact scorecard (Fabre et al. 2008) or performance surface idea are examples of how to go from various acoustic and environmental factors to a measure that is operationally meaningful. There are other approaches as well. The committee should be able to define a range of approaches, building on work done in the studies briefed at the workshop and elsewhere and put this together as a “code of best practice” for estimating METOC impacts on warfighting effectiveness. DoD and NATO have developed something similar called the “C4I Analysis Code of Best Practice” ([www.dodccrp.org](http://www.dodccrp.org)).

The next step in this metrics process will then be to research, identify, and propose an approach or approaches for development of the aforementioned methodology. This approach will be different for various systems, but the initial focus will be on the current ASW systems discussed during the workshop. An example approach would be to develop a generic scenario for which environmental acoustic products (e.g., Performance Surface) and their variations can be applied to quantify the operational impacts of the product in terms that can be communicated to decision makers. Stevens et al. (2008) provide examples that can be followed for this type of approach.

It is expected that the scientists that participated in this workshop will continue to develop and improve their metrics for more easy translation to higher level metrics. It is well understood that there is no “silver bullet” metric, however, steps can be and are being made to bring the two communities closer together.

Another Technical Metrics Workshop is tentatively planned for FY10. The purpose of that workshop will be threefold:

1. to present new technical metrics and progress on existing technical and related metrics since the 2008 workshop
2. to develop and refine the general procedure for deriving operational metrics from technical metrics; document the issues involved; and potentially to begin applying the procedure to a test case; and
3. to get feedback from the various entities involved on the technical metrics way ahead.

**ACKNOWLEDGMENTS**

The author would like to thank the committee (Emanuel Coelho, NRL SSC Code 7320 (oceanography)/University of Southern Mississippi (USM), James Dykes, NRL SSC Code 7320 (oceanography), Dr. Pat Gallacher, NRL SSC Code 7330 (oceanography), Dr. Roger Gauss, NRL Code 7140 (acoustics), Joe Metzger, NRL SSC Code 7320 (oceanography) and Dr. Tom Murphree, NPS (meteorology/ oceanography/ climate/metrics)) for their hard work towards the success of the workshop and to the committee and Dr. Bill Stevens for their significant contribution to this report; it was definitely a group effort.

Thank you to the Naval Research Laboratory Code 7000 and the Office of Naval Research for encouraging and funding this effort.